Tentamentsskrivning i **Matematisk Statistik med Metoder MVE490**

Tid: den 22 december, 2016
Examinatorer: Kerstin Wiklander och Erik Broman.
Jour: Kerstin Wiklander (tel 031-772 5355).
Hjälpmedel: miniräknare, egenhändigt skriven formelsamling om två A4 fram och bak (dvs 4 sidor) samt tabeller (delas ut på plats).

Tentamen består av 8 frågor om sammanlagt 50 poäng. Preliminära betygsgränser är satta till:
betyg "3": 20 till 29
betyg "4": 30 till 39 poäng
betyg "5": 40 eller fler poäng.

OBS! Alla lösningar skall vara väl redovisade och motiverade. Talen är ej ordnade efter svårighetsgrad

1. Hjälp Mimmi att svara på följande frågor. (5p)

   (a) (2p) Beräkna median, medelvärde, standardavvikelse, och SE"(standard error of the mean, dvs medelvärdets spridningsmått) från följande datamängd: 12.5 18.4 17.5 11.9 15.2

   (b) (1p) Vilket av påståendena nedan om en nollhypotes är korrekt?
   Alternativ 1) Den betecknas med H_a (eller H_1).
   Alternativ 2) Den är alltid fel med sannolikheten 5%.
   Alternativ 3) Den är till hjälp för att dra slutsatsen att t ex en åtgärd inte har någon effekt.
   Alternativ 4) Den uttrycks alltid med hjälp av någon eller några parametrar.
   Alternativ 5) Alla påståenden 1-4 ovan är korrekta.

   (c) (2p) Antag att vi har en normalfördelningsmodell. Vi har dessutom en mothypotes $H_a : \mu < \mu_0$ där μ_0 är något värde enligt nollhypotesen. Om man använder en teststatistika där man vet vad σ^2 är, vad blir gränsen till förkastelseområdet då signifikansnivån är vald till 1%?
   Om man istället inte känner till σ^2 och utnyttjar data med stickprovsstorleken $n = 9$, vad blir gränsen till förkastelseområdet då?

   Lösning:

   a) Medianen=15.2, $\bar{x} = 15.1$, s=2.9, sem=$s/\sqrt{5}$=1.3.
   b) Alternativ 4)
   c) Enligt bokens sätt att beteckna percentiler (se sid 316), dvs som $P(Z > z_\alpha) = \alpha$ så är gränserna till förkastelseområdena $z_{0.99} = -z_{0.01} = -2.326$ och $t_{8,0.99} = -t_{8,0.01} = -2.896$.

2. Betrakta två händelser A, B och antag att $\mathbb{P}(A) = 0.2$ och $\mathbb{P}(B) = 0.7$ (6p)
   (a) Vad blir $\mathbb{P}(A \cup B)$ och $\mathbb{P}(A^c \cup B)$ om A, B är disjunkta?
   (b) Vad blir $\mathbb{P}(A \cup B)$ och $\mathbb{P}(A^c \cup B)$ om A, B är oberoende?
   (c) Vad blir $\mathbb{P}(A|B)$ och $\mathbb{P}(A^c|B)$ om $\mathbb{P}(A \cap B) = 0.1$?

   Lösning:

   (a) Vi har att
   $$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.2 + 0.7 = 0.9.$$
   Om A, B är disjunkta så måste $B \subset A^c$. Därför får vi att
   $$\mathbb{P}(A^c \cup B) = \mathbb{P}(A^c) = 1 - 0.2 = 0.8.$$

   (b) Om A, B är oberoende så har vi att
   $$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = 0.2 + 0.7 - 0.2 * 0.7 = 0.76.\end{aligned}$$
   På liknande sätt får vi att
   $$\begin{aligned}\mathbb{P}(A^c \cup B) &= \mathbb{P}(A^c) + \mathbb{P}(B) - \mathbb{P}(A^c \cap B)\\ &= \mathbb{P}(A^c) + \mathbb{P}(B) - \mathbb{P}(A^c)\mathbb{P}(B) = 0.8 + 0.7 - 0.8 * 0.7 = 0.94.\end{aligned}$$

   (c) Vi har att
   $$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.1}{0.7} = \frac{1}{7},$$
   och att
   $$\mathbb{P}(A^c|B) = \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B) - \mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.7 - 0.1}{0.7} = \frac{6}{7}.$$

3. Harry spelar tärning. Om Harry slår 1,2 eller 3 förlorar han 100 SEK. Om Harry slår en 4 eller 5 får han 100 SEK, och om han slår en 6 får han 200 SEK. (8p)
   (a) Om Harry har 500 SEK, vad är den förväntade summan pengar Harry har efter ett spel?
   (b) Låt X vara antalet gånger som Harry får 6 om han spelar 100 gånger. Vilken fördelning har X? Motivera.
   (c) Låt Y vara antalet gånger som Harry får 4 eller 5 om han spelar 100 gånger. Vilken fördelning har Y? Motivera.
   (d) Vad är sannolikheten att Harry slår 6 fler än 24 gånger om han kastar tärningen 100 gånger?

Lösning:

(a) Låt V vara förtjänsten för ett slag och T resultatet av tärningsslaget. Vi har då att

$$\begin{aligned}\mathbb{E}[V] &= -100\mathbb{P}(T=1) - 100\mathbb{P}(T=2) - 100\mathbb{P}(T=3) \\ &\quad +100\mathbb{P}(T=4) + 100\mathbb{P}(T=5) + 200\mathbb{P}(T=6) \\ &= -3\frac{100}{6} + 2\frac{100}{6} + \frac{200}{6} = \frac{100}{6}.\end{aligned}$$

Den förväntade summan är därför $500 + 100/6$.

(b) Då vi har oberoende experiment som lyckas med sannolikhet $1/6$ har vi att $X \sim$Bin(100,1/6).

(c) Då vi har oberoende experiment som lyckas med sannolikhet $2/6$ har vi att $X \sim$Bin(100,2/6).

(d) Vi har att $\mathbb{E}[X] = 100/6$ och $Var(X) = 500/36$

$$\mathbb{P}(X \geq 24) = \mathbb{P}\left(\frac{X - \frac{100}{6}}{\sqrt{\frac{500}{36}}} \geq \frac{24 - \frac{100}{6}}{\sqrt{\frac{500}{36}}}\right) \approx \mathbb{P}(Z \geq 1.9677) \approx 0.975$$

4. Glennita skall renovera sin fasad och köper därför material hos RekoBygg AB. Hon köper en väldig massa spik (2000) varav ett fåtal innehåller defekter. Glennita antar att antalet defekta spikar är Poissonfördelade med någon parameter λ. (6p)

(a) Är Glennitas antagande om Poissonfördelning rimlig? Vad talar för? Vad talar emot?

(b) Om $\lambda = 5$, vad är sannolikheten att minst tre spikar är defekta?

(c) Vad är sannolikheten att exakt fyra spikar är defekta om Glennita vet att minst tre är defekta?

Lösning:

(a) Ja och Nej. Om man antar att tillverkningsprocessen producerar defekta spikar oberoende av varandra, så är den exakta fördelningen Bin$(2000, p)$ där $p =$sann för en defekt. Om p är litet så är Poisson en lämplig approximativ fördelning. Dock skulle man med Poissonfördelningen kunna få fram att antalet defekta spikar är fler än 2000, vilket såklart är orimligt.

(b) Låt $X \sim$Poi(5). Vi söker då

$$\begin{aligned}&\mathbb{P}(X \geq 3) \\ &\quad = 1 - \mathbb{P}(X \leq 2) = 1 - \mathbb{P}(X=0) - \mathbb{P}(X=1) - \mathbb{P}(X=2) \\ &\quad = 1 - e^{-5}\left(1 + 5 + \frac{5^2}{2}\right) \approx 0.8753\end{aligned}$$

(c) Vi söker

$$\mathbb{P}(X = 4|X \geq 3) = \frac{\mathbb{P}(X = 4, X \geq 3)}{\mathbb{P}(X \geq 3)}$$
$$= \frac{\mathbb{P}(X = 4)}{1 - \mathbb{P}(X \leq 2)} = \frac{e^{-5}\frac{5^4}{4!}}{1 - e^{-5}\left(1 + 5 + \frac{5^2}{2}\right)} \approx 0.2005$$

5. Bert spelar poker. En pokerhand består av fem kort (där ordningen som korten kommer i inte spelar någon roll). (4p)

   (a) Hur många pokerhänder finns det totalt?

   (b) Hur många pokerhänder finns det som inte innehåller några spader?

   (c) Vad är sannolikheten att Bert får färg (dvs alla fem kort har samma färg) på given, dvs innan eventuella byten av kort genomförs?

   Lösning:

   (a) Det finns $\binom{52}{5}$ händer totalt.

   (b) Om vi tar bort alla spader finns det 39 kort kvar. Svaret blir då $\binom{39}{5}$.

   (c) Det finns 13 hjäret, så antalet händer med bara hjärter är $\binom{13}{5}$. Total finns det därför $4\binom{13}{5}$ händer med färg, och däför blir sannolikheten

$$\frac{4\binom{13}{5}}{\binom{52}{5}} \approx 0.002$$

6. Kajsa jobbar med PR i ett företag som har ett stort antal butiker. En reklamkampanj gjordes för att öka försäljningsintäkterna. Förändringen för dessa från sju slumpmässigt utvalda butiker blev (i tusen kronor): 11.4 21.5 −17.2 17.6 47.2 −29.1 41.1

   Kajsa ger dig i uppdrag att testa på signifikansnivån 5% om reklamkampanjen har varit lyckosam för företaget. Introducera variabel, ange modell och hypoteser. Vad blir slutsatsen? Skriv den också i klartext som kan användas i en rapport till Kajsa. (9p)

   Lösning:

   Låt X = Förändringarna i försäljningssiffrorna. Modell: $X \sim N(\mu, \sigma^2)$. $H_0 : \mu = 0$, $H_a : \mu > 0$. Gräns till förkastelseområdet: $t_{6,0.05} = 1.943$. Teststatistikans värde blev 1.244. Eftersom teststatistikan inte hamnat i förkastelseområdet kan H_0 inte förkastas på signifikansnivå 5%. Man kan inte i denna undersökning påvisa att reklamkampanjen har ökat de förväntade försäljningsintäkterna.

7. Nisse har undersökt två olika leverantörer av julklappar för att sedan välja den bästa när det gäller leveranstider. Dessa (i timmar) registrerades för ett antal körningar med jämförbara produkter och leveransadresser. Nisse

anser att de två variablerna är oberoende och följer normalfördelningen med samma teoretiska varians. Syftet var att undersöka om deras förväntade leveranstider skiljer sig åt.
Sammanfattande mått från undersökningen ges nedan. (7p)

	Medelvärde	Standardavvikelse (s)	Stickprovsstorlek
Leverantör 1:	2.23	0.31	10
Leverantör 2:	1.78	0.32	9

(a) (2p) Formulera hypoteserna.

(b) (3p) Beräkna ett 99%-igt konfidensintervall för skillnaden mellan de två väntevärdena.

(c) (2p) Visa hur du kan använda det uträknade konfidensintervallet som enda grund till att testa nollhypotsen i a)-uppgiften. Ange också vad signifikansnivån blir i ett sådant test.

Lösning:

a) $H_0 : \mu_1 = \mu_2$, $H_a : \mu_1 \neq \mu_2$.
b) Om Nisses antaganden är rätt beräknas konfidensintervallet med värden ur t-fördelningen. Värdet på det 99%-iga konfidensintervallet för $\mu_1 - \mu_2$ blir $(0.044, 0.856)$.
c) Eftersom nollan inte ingår i konfidensintervallet kan H_0 förkastas. Signifikansnivå är 1% eftersom konfidensgraden är 99%.

8. Kalle har tagit reda på att 28% av hushållen i ett visst bostadsområde har fler än en bil. Därefter gjorde han en enkätundersökning bland slumpmässigt utvalda hushåll i ett nytt bostadsområde i närheten. Svaren han fick in blev att 34 bland de 164 utvalda hushållen hade fler än en bil.

   Hjälp Kalle att testa på signifikansnivån 5% om andelen hushåll med fler än en bil i det nya området avviker från det närliggande bostadsområdets. Glöm inte att formulera hypoteserna. (4p)

   Lösning:

   Hypoteserna: $H_0 : p = 0.28$, $H_a : p \neq 0.28$. Punktskattning av proportionen p i det nya området är $\hat{p} = 34/164 \approx 0.207$. Förutsättningarna för approximation med normalfördelningen är uppfyllda (se i boken sid 333). Den teoretiska variansen för $\hat{p}$ är $\frac{p(1-p)}{n}$. Skatta den genom att ersätta p med p_0, där p_0 är värdet under nollhypotesen $H_0 : p = p_0$ (här är p_0= 0.8). Teststatistikans värde blir $(\hat{p} - p_0)/\sqrt{(p_0 - (1-p_0)/n} = (0.207 - 0.28)/\sqrt{(0.28(1-0.28)/164} \approx -2.07$.

   Jämför detta med percentilernas värden $\pm z_{0.025} = \pm 1.96$ (gränser till förkastelseområdet). Eftersom teststatistikan har hamnat i förkastelseområdet kan H_0 förkastas på signifikansnivå 5%. Man kan påvisa att andelen hushåll med fler än två bilar i det nya området skiljer sig åt från det närliggande bostadsområdet. Det visade sig att det är en signifikant lägre andel i det nya området.